Weekly Report

Period: 2016/7/4-2016/7/10
Reporter: Li Zongzhuang

Node, Node-Link, and Node-Link-Group Diagrams: An Evaluation

(from TVCG 2014 ,Author Bahador Saket and so on)

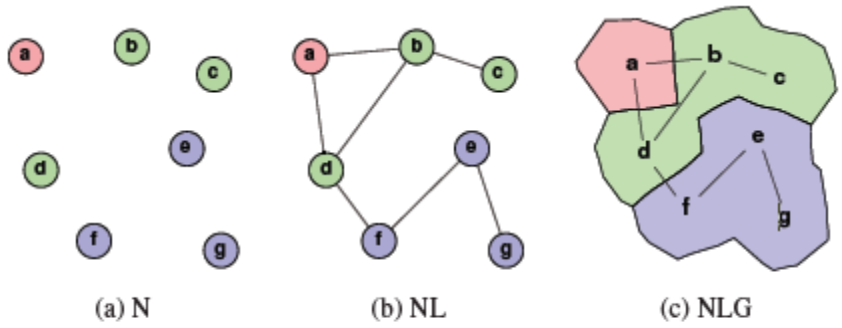


Figure 1: Examples of diagrams considered in this study.

There are three Techniques: node diagrams, node-link diagrams and node-link-group diagrams.

This paper assesses these three types of diagrams with a controlled experiment that covers nine different tasks falling broadly in three categories: node-based tasks, network-based tasks and group-based tasks.

Different ways are used like size and density and color selection. Authors used three different datasets.

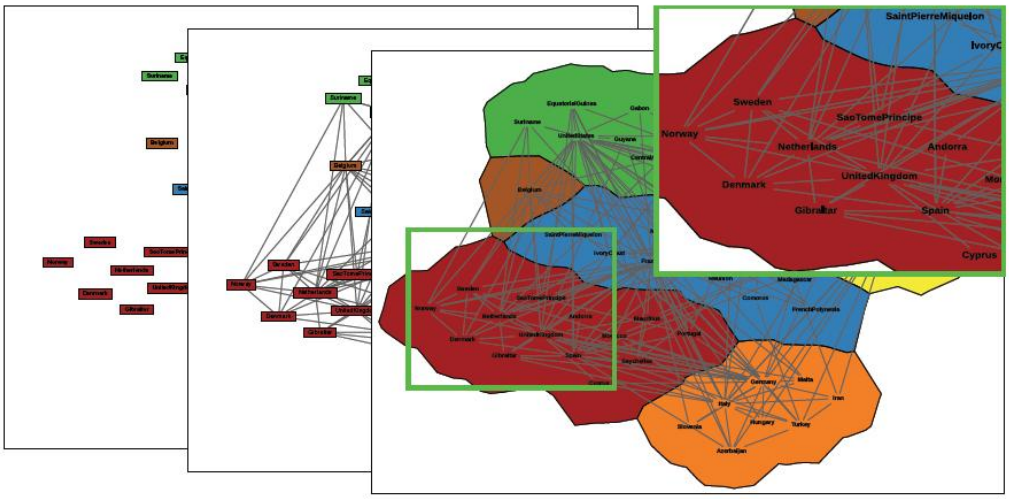


Figure 5: Representation of 50 nodes and 200 links with N, NL and NLG diagrams; underlying data from the World-trade dataset.

NLG diagrams outperform N and NL diagrams, but this remains to be studied. As there are different node-link-group visualizations, more work is also needed to evaluate different NLG diagram generation methods such as Bubblesets, Linesets, Kelp diagrams, and GMap.

Such results can have significant implications for the design of standard and domain specific visualizations tools.

Biperpedia: An Ontology for Search Applications **(Author** Rahul Gupta and so on)

Biperpedia, an ontology with 1.6M (class, attribute) pairs and 67K distinct attribute names. Biperpedia extracts attributes from the query stream, and then uses the best extractions to seed attribute extraction from text.

First, they develop methods for classifying different relationships between pairs and attributes and discovering them from Web text. Second, They mine a grammar for complex attribute names.

This paper also shows that it can increase the number of Web tables whose semantics we can recover by more than a factor of 4 compared with Freebase.

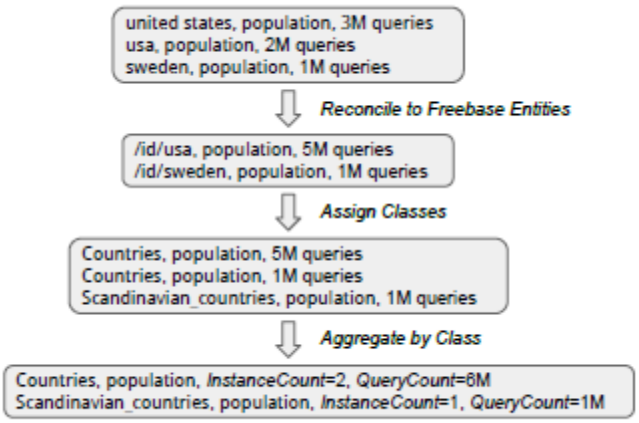
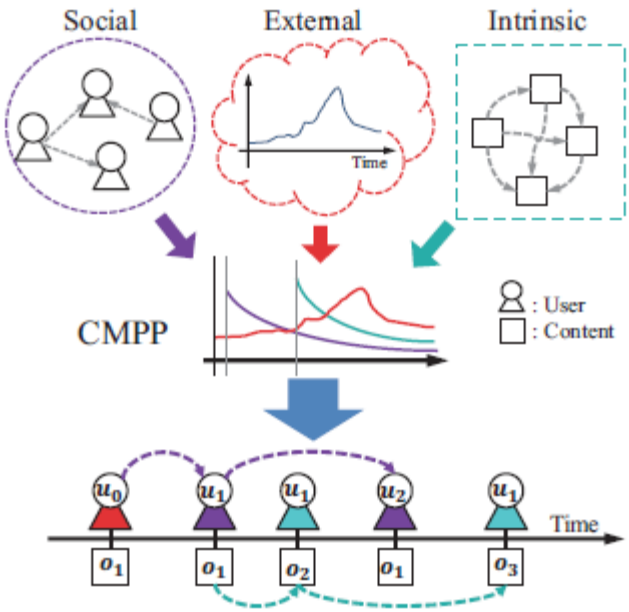


Figure 3: Query stream extraction

Why It Happened: Identifying and Modeling the Reasons of the Happening of Social Events **(from ACM ,Author** Yu Rong and so on)

Combinational Mixed Poisson Process (CMPP) model, can explain not only how information diffuses in social networks, but also why a specific event happens.

Authors use the mixed Poisson process to model event cascade generated by different factors respectively and integrate different Poisson processes with shared parameters. They distinguish three factors: social, external and intrinsic influence which can explain the emergence of every specific event.



Matches, Mismatches, and Methods: Multiple-View Workflows for Energy Portfolio Analysis **(from TVCG, Author** Tamara and so on)

The main contribution of this paper is a methodological advice for visualization design projects, which includes considerations for designing workflows that incorporate multiple views. This energy analysis workflow can also be reflected in the task abstractions including overview, drill down and roll up.

Instant Espresso: Interactive Analysis of Relationships in Knowledge Graphs

(from ACM**, Author Stephan Seufert** and so on)

This paper shows InstantEspresso, a system for interactive analysis of relationships between two sets user-specified entities over a knowledge graph.

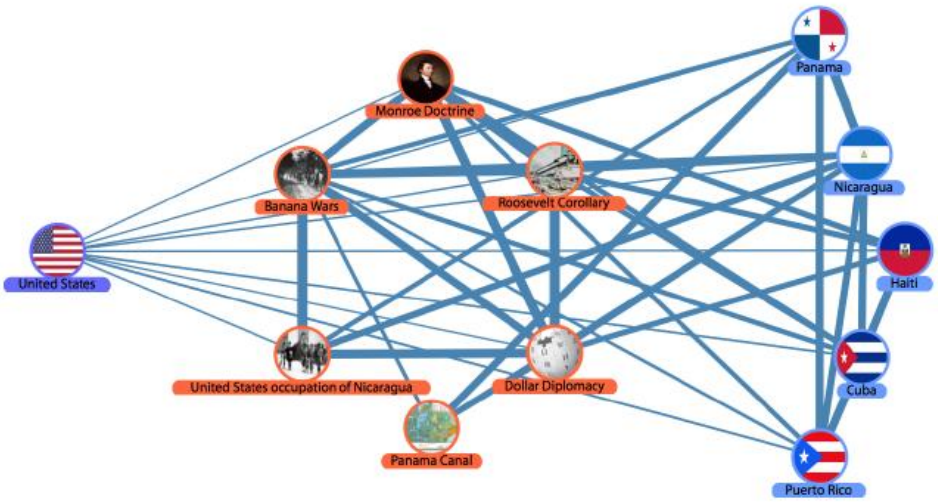


Figure 3: Relatedness Core between the United States and countries from the Americas

Espresso knowledge graph and relatedness core computation are used to compute the results and relatedness cores.

The paper also provides some use cases, such as graphical relationship summarization aggregated relationship summarization retrieval of relevant documents.

Knowledge Graph and Visualization

Li Zongzhuang

Abstract: The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. The Knowledge Graph uses the power of semantics, and wants to improve search precision and effectiveness by building the semantic web. Visualization is the study of (interactive) visual representations of abstract data to reinforce human cognition. The visualization of Knowledge Graph can Effectively improve the efficiency of the user to complete the search target and precision. Data integration and correlation analysis reasoning is one of the best visual analysis applications. At this time, There are many applications have been exploited based on this concept.

1. Introduction

Knowledge graphs on the Web are a backbone of many information systems that require

access to structured knowledge. The idea of feeding intelligent systems and agents with general, formalized knowledge of the world dates back to classic Artificial Intelligence research in the 1980s. Then, with the advent of Linked Open Data sources like DBpedia, and by Google's announcement of the Google Knowledge Graph in 2012, representations of general world knowledge as graphs draw a lot of attention again.

Google's Knowledge Graph display was added to Google's search engine in 2012. Once a user search one thing, it provides structured and detailed information about the topic in addition to a list of links to other sites. According to Google, the information in the Knowledge Graph is derived from many sources, including the CIA World Factbook, Wikidata, and Wikipedia.

In 2014, Google announced a new initiative,called the Knowledge Vault, which derives much of its data from the Knowledge Graph and the sources thereof, as well as harvesting its own data, ranking its reliability and compiling all results into a database of over 1.6 billion facts collected by machine learning algorithms.

In a paper , author shows what is a knowledge graph:

1. mainly describes real world entities and their interrelations, organized in a graph.
2. defines possible classes and relations of entities in a schema.
3. allows for potentially interrelating arbitrary entities with each other.
4. covers various topical domains.

Vision is the most important channels to the information of the outside world. Visualization is the data technology of interactive visual expression. On the century of big data, The ability of processing data is far behind the ability to get the data. The amount of data contained in Knowledge Graph is huge, so the visualization can be an important means of Knowledge Graph data processing. It can help us find the phenomena and laws faster and achieve the goal. However, the research about the visualization of Knowledge Graph is relatively shallow.

2. Knowledge Graph

2.1 Knowledge Graphs

There are many ways to build knowledge graphs. They can be curated like *Cyc*, edited by the crowd like *Freebase* and *Wikidata*, They can also be extracted from large-scale, semi-structured web knowledge bases such as Wikipedia, *DBpedia* and *YAGO*. Furthermore, information extraction methods for unstructured or semi-structured information are proposed, which lead to knowledge graphs like *NELL*, *PROSPERA*, or *KnowledgeVault*.

Name	Instances	Facts	Types	Relations
DBpedia (English)	4,806,150	176,043,129	735	2,813
YAGO	4,595,906	25,946,870	488,469	77
Freebase	49,947,845	3,041,722,635	26,507	37,781
Wikidata	15,602,060	65,993,797	23,157	1,673
NELL	2,006,896	432,845	285	425
OpenCyc	118,499	2,413,894	45,153	18,526
Google's Knowledge Graph	570,000,000	18,000,000,000	1,500	35,000
Google's Knowledge Vault	45,000,000	271,000,000	1,100	4,469
Yahoo! Knowledge Graph	3,443,743	1,391,054,990	250	800

An overview about these knowledge graphs.

2.2 Semantic Web

The core concept of Knowledge is the introduction of the semantic, which means that let the computers know the semantic judgments.

3. Visualization

3.1 Graph data visualization

Graph data is an important component in data. The visualizations about graph data are often presented by node-link graph.

3.2 High-dimensional data visualization

There are many data sets have more than one dimension. So many visualization tools have been invented to present high-dimensional data. The results got by Knowledge Graph often have many properties. That means we can get some revelation.

3.3 Another types visualization

Because of the difference of goals, there can be many visualization schemes. Maybe we can learn more from them.

4. Application softwares

Based on the theory of human-computer interaction, there are a lot of software is put forward based on the semantic web. This kind of software focus on data integration and correlation analysis. Data integration made in background automatically, and data correlation analysis mainly rely on people's reasoning ability as well as front end some interactions. That’s the best application, which give full play to the people the calculation of analytical reasoning skills and computer specialty.

There are many applications in this area, such palantir, IBM i2, Tableau and so on.